

## A Suggested Nomenclature for Bacterial Host Modification and Restriction Systems and their Enzymes

In the proposed nomenclature restriction–modification systems are named according to host organism and strain. Different R–M† systems in a single host are designated by Roman numerals. Restriction nucleases and modification methylases are given the general names endonuclease R and methylase M, followed by their R–M system name.

Since the original report of the isolation of a restriction endonuclease from *Escherichia coli* K by Meselson & Yuan (1968) a number of such enzymes have been purified from various strains of bacteria including several that are specified by intracellular viruses (Meselson & Yuan, 1968; Haberman *et al.*, 1972) or plasmids (Yoshimori, 1971). Restriction endonucleases from *Hemophilus* (Smith & Wilcox, 1970; Gromkova & Goodgal, 1972; Middleton *et al.*, 1972) and from *E. coli* carrying resistance transfer factors (Yoshimori, 1971) produce site-specific cleavage of DNA (Kelly & Smith, 1970; Hedgpeth *et al.*, 1972; Boyer *et al.*, 1973) and are proving increasingly useful as tools for DNA and chromosomal analysis (Danna & Nathans, 1971; Middleton *et al.*, 1972; Mertz & Davis, 1972; Morrow & Berg, 1972; Mulder & Delius, 1972; Edgell *et al.*, 1972; Johnson *et al.*, 1973). This has led to a search for additional enzymes. It seems advisable to introduce general rules for naming each new enzyme now, rather than after a large number have come into use carrying arbitrary names. Because of the variety of organisms from which they are isolated some kind of species–strain designation for each new enzyme would appear to be reasonable. A system of nomenclature for DNA fragments produced by restriction endonucleases and DNA sites cleaved by such enzymes has been suggested by Danna *et al.* (1973). This system utilizes a three-letter abbreviation for the genus–species name of the bacterial host and could provide the basis for a more general nomenclature.

A restriction enzyme is a component of a restriction–modification system of a given specificity within an organism. The R–M† system consists of two enzymatic components, a restriction endonuclease, and a modification enzyme sharing similar (or identical) recognition specificity. We would like to propose that the genus–species designation be used as an R–M system name according to the following rules.

(1) The genus and species name of the host organism is identified by the first letter of the genus and the first two letters of the species to form a three-letter abbreviation in italics. For example: *E. coli*, *Eco* and *H. influenzae*, *Hin*.

(2) Strain or type identification follows the genus–species abbreviation in non-italicized symbols, e.g. *EcoB* or *EcoK*. In cases where the R–M system is genetically specified by a virus or plasmid the italicized genus–species abbreviation of the host is given and the symbol for the extrachromosomal element follows in non-italics, e.g. *EcoP1*, *EcoRI*, etc. In occasional cases where it might be necessary to specify

† Abbreviations used: R, restriction; M, modification.

TABLE I  
A listing of some restriction-modification systems

Host	Restriction-modification system	How identified <sup>1</sup>	Modification site	Restriction site	Reference <sup>5</sup>
<i>H. parainfluenzae</i>	<i>Hpa</i> I	R <sup>2</sup>			1, 2
	<i>Hpa</i> II	R			
<i>H. aegyptius</i>	<i>Hae</i>	R		5' GG ↓ CC CC ↑ GG 5'	3, 26
<i>H. influenzae</i>	<i>Hina</i> I	G			4
	<i>Hina</i> II	G			
	<i>Hin</i> b	G			4
	<i>Hinc</i> II	R			5
	<i>Hind</i> I	G, M	5' C <sup>*</sup> AC <sup>3</sup>		6, 7
	<i>Hind</i> II	M, R	5' Pu <sup>*</sup> AC	5' GTPy ↓ Pu <sup>*</sup> AC CAPu ↑ PyTG 5' *	8, 9, 7
	<i>Hind</i> III	M, R	5' N <sup>*</sup> AA	5' A ↓ AGCTT T TCGA ↑ A 5'	7, 26
	<i>Hine</i>	G			4
	<i>Hinf</i>	G			4
<i>E. coli</i>			Contains dinucleotides		
	<i>Eco</i> B	G, M, R	5' GA <sup>*</sup> , CA <sup>*</sup> , AA <sup>*</sup> , AC <sup>*</sup>		10, 11, 12, 13
	<i>Eco</i> K	G, M, R			10, 14, 15
	<i>Eco</i> 15	G			16

	<i>EcoA</i>	G			17
	<i>EcoRI</i> <sup>1</sup>	G, M, R	5' A <sup>*</sup> AT	5' (A/T)G ↓ A <sup>*</sup> ATT C(T/A) (T/A)C TTAA ↑ G(A/T) 5'	18, 19, 20
	<i>EcoRII</i>	G, M, R	5' C <sup>*</sup> CT	5' ↓ C <sup>*</sup> CTGG GGACC ↑ 5'	18, 21, 22
			5' CCA		
	<i>EcoP1</i>	G, M, R	5' AGA <sup>*</sup> TCT		17, 23, 15, 14, 26
<i>S. typhimurium</i>	<i>StyLT</i>	G			24
	<i>StyN3</i>	G			25

<sup>1</sup> A host R-M system may be detected in an organism by biological or biochemical means. By biological tests, strain Y carries such a system if the following holds: (1) a virus grown in some strain X plates with reduced efficiency on strain Y relative to X, i.e. growth is "restricted" by Y; (2) virus recovered from strain Y plates with full efficiency on strain Y, i.e. is "modified" by Y, but if regrown on strain X, again plates with reduced efficiency on strain Y. These are the original criteria of Luria & Human (1952). The biochemical basis for these biological phenomena are now relatively well understood (for reviews see Arber & Linn, 1969; Boyer, 1971; Meselson *et al.*, 1972). Phage grown on strain X is restricted on Y because the viral DNA does not have the modification pattern of strain Y. DNAs lacking appropriate modification are recognized by a site-specific restriction endonuclease and double-stranded cleavage of the infecting viral DNA results. Viral DNA occasionally escapes cleavage and gives rise to progeny virus, which receives Y-specific modifications. The virus then plates with full efficiency on strain Y. Regrowth on strain X, however, results in one-cycle loss of Y-specific modification. Modification is produced in most cases by a DNA methylase having site-specific recognition similar or identical to its corresponding restriction endonuclease partner.

Recently, several restriction endonucleases have been detected in strains of *Hemophilus* by purely biochemical assays. These are identified as restriction enzymes because they recognize and cleave specific short oligonucleotide sequences in foreign DNA while remaining inactive on the presumably modified host DNA. A rigorous biochemical definition of a host modification and restriction system requires the following: (1) the endonuclease recognizes a specific oligonucleotide sequence on DNA and (2) as a result produces cleavage which may be within the recognition site or at some distance along the DNA; (3) no cleavage is produced in the absence of recognition sites; (4) specific modification of the recognition sites prevents recognition and cleavage by the restriction enzyme.

<sup>2</sup> R, restriction endonuclease identified; G, identified by efficiency of plating of phage; M, modification methylase identified.

<sup>3</sup> Asterisks in this and the following sequences indicate the site of methylation by the corresponding methylase, M.

<sup>4</sup> The use of RI and RII instead of the plasmid names RTF-1 and RTF-2 seems advisable here because of the extensive association of these symbols with the restriction enzymes from the two plasmids.

<sup>5</sup> 1, Gromkova & Goodgal, 1972; 2, Sharp *et al.*, 1973; 3, Middleton *et al.*, 1972; 4, Piekarowicz & Glover, 1972; 5, Landy, A., personal communication; 6, Glover & Piekarowicz, 1972; 7, Roy & Smith, 1973a,b; 8, Smith & Wilcox, 1970; 9, Kelly & Smith, 1970; 10, Arber, 1965; 11, Kühnlein & Arber, 1972; 12, Linn & Arber, 1968; 13, vanOrmondt, personal communication; 14, Meselson & Yuan, 1968; 15, Haberman *et al.*, 1972; 16, Arber & Wauters-Williams, 1970; 17, Arber *et al.*, 1972; 18, Yoshimori, 1971; 19, Hedgpeth *et al.*, 1972; 20, Boyer, personal communication; 21, Bigger *et al.*, 1973; 22, Boyer *et al.*, 1973; 23, Brockes *et al.*, 1972; 24, Colson & Colson, 1967; 25, Hattman, 1971; 26, Murray, K., personal communication.

the host strain as well as the extrachromosomal element the strain identification symbol may be inserted parenthetically, e.g. *Eco*(B)P1.

(3) When a particular host strain has several different R-M systems, these are identified by Roman numerals, thus, the R-M systems from *H. influenzae* strain d would be *Hind*I, *Hind*II, *Hind*III, etc.

All restriction enzymes could then have the general name endonuclease R, but in addition carry the system name, e.g. endonuclease R·*Eco*B. Similarly, modification enzymes could be named according to the modifying group introduced, e.g. methylase M or glucosylase M followed by the system name. The modification enzyme from *E. coli* B corresponding to endonuclease R·*Eco*B would then be designated methylase M·*Eco*B. These enzyme names may be further shortened to endo R·*Eco*B and meth M·*Eco*B.

In many cases, where only one type of organism is under discussion, the genus-species abbreviation can be dropped from the R-M system name leaving only the strain and Roman numeral designation. For example, a paper dealing with R-M systems in *H. influenzae* strains a and d would use the names aI, aII, dI, dII, etc. Enzymes would then be named endo R·aI, etc.

Occasionally the situation may arise where a restriction enzyme has been given a Roman numeral designation, e.g. endo R·dI, but subsequently is proved to consist of two enzymes of different specificity belonging to different R-M systems. These may be differentiated as endo R·dIa and endo R·dIb.

Table 1 is a listing of some R-M systems, illustrating the suggested genus-species nomenclature, and indicating the nucleotide sequence specificity of the system where it is known. Since it is illustrative no attempt at completeness of either the list or the references is made. The rules given here for R-M systems nomenclature are compatible with, and conveniently used with, the genetic nomenclature recently proposed by Arber (personal communication).

In regard to nomenclature for DNA fragments produced by restriction endonucleases and the sites of cleavage within a DNA molecule, we suggest that fragments be designated by capital letters in order of decreasing size and cleavage sites be designated by arabic numerals sequentially along the molecule. To specify a particular endonuclease the letter or number would be preceded by the R-M system name suggested above. For example, endonuclease R·*Hind*III produces six fragments from SV40 DNA designated *Hind*III-A, B, etc. In most cases this notation can be further abbreviated to dIII-A, B, etc.

This proposal was circulated prior to publication and its final form is the result of suggestions and contributions from: W. Arber, P. Berg, H. W. Boyer, S. W. Glover, S. Hattman, T. J. Kelly Jr, A. Landy, S. Linn, K. Murray, P. H. Roy, H. vanOrmondt, J. Sambrook, R. L. Sinsheimer, S. M. Weissman and K. W. Wilcox.

Department of Microbiology  
The Johns Hopkins University  
School of Medicine  
Baltimore, Md 21205, U.S.A.

HAMILTON O. SMITH  
DANIEL NATHANS

Received 10 September 1973

## REFERENCES

- Arber, W. (1965). *Annu. Rev. Microbiol.* **19**, 365-378.
- Arber, W. & Linn, S. (1969). *Annu. Rev. Biochem.* **38**, 467-500.
- Arber, W. & Wauters-Williams, D. (1970). *Mol. Gen. Genet.* **108**, 203-217.
- Arber, W., Rifat, A., Wauters-Williams, D. & Kühnlein, U. (1972). *Mol. Gen. Genet.* **115**, 195-207.
- Bigger, C. H., Murray, K. & Murray, N. E. (1973). *Nature New Biol.* **244**, 7-10.
- Boyer, H. W. (1971). *Annu. Rev. Microbiol.* **25**, 153-176.
- Boyer, H. W., Chow, L. T., Dugaiczky, A., Hedgpeth, J. & Goodman, H. M. (1973). *Nature New Biol.* **244**, 40-43.
- Brookes, J., Brown, P. R. & Murray, K. (1972). *Biochem. J.* **127**, 1-10.
- Colson, C. & Colson, A. M. (1967). *Biochem. Biophys. Res. Commun.* **29**, 692-695.
- Danna, K. J. & Nathans, D. (1971). *Proc. Nat. Acad. Sci., U.S.A.* **68**, 2913-2917.
- Danna, K. J., Sack, G. & Nathans, D. (1973). *J. Mol. Biol.* **78**, 363-376.
- Edgell, M. H., Hutchison, C. A., III & Sclair, M. (1972). *J. Virol.* **9**, 574-582.
- Glover, S. W. & Piekarczyk, A. (1972). *Biochem. Biophys. Res. Commun.* **46**, 1610-1617.
- Gromkova, R. & Goodgal, S. H. (1972). *J. Bacteriol.* **109**, 987-992.
- Haberman, A., Heywood, M. & Meselson, M. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3138-3141.
- Hattman, S. (1971). *J. Virol.* **7**, 690-691.
- Hedgpeth, J., Goodman, H. M. & Boyer, H. W. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3448-3452.
- Johnson, P. H., Lee, A. S. & Sinsheimer, R. L. (1973). *J. Virol.* **11**, 596-599.
- Kelly, T. J. & Smith, H. O. (1970). *J. Mol. Biol.* **51**, 393-409.
- Kühnlein, U. & Arber, W. (1972). *J. Mol. Biol.* **63**, 9-16.
- Linn, S. & Arber, W. (1968). *Proc. Nat. Acad. Sci., U.S.A.* **59**, 1300-1306.
- Luria, S. E. & Human, M. L. (1952). *J. Bacteriol.* **64**, 557-569.
- Mertz, J. E. & Davis, R. W. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3370-3374.
- Meselson, M. & Yuan, R. (1968). *Nature (London)*, **217**, 1110-1114.
- Meselson, M., Yuan, R. & Heywood, J. (1972). *Annu. Rev. Biochem.* **41**, 447-466.
- Middleton, J. H., Edgell, M. H. & Hutchison, C. A., III (1972). *J. Virol.* **10**, 42-50.
- Morrow, J. F. & Berg, P. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3365-3369.
- Mulder, C. & Delius, H. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3215-3219.
- Piekarczyk, A. & Glover, S. W. (1972). *Mol. Gen. Genet.* **116**, 11-25.
- Roy, P. H. & Smith, H. O. (1973a). *J. Mol. Biol.* In the press.
- Roy, P. H. & Smith, H. O. (1973b). *J. Mol. Biol.* In the press.
- Sharp, P. A., Sugden, B. & Sambrook, J. (1973). *Biochemistry*, **12**, 3055-3063.
- Smith, H. O. & Wilcox, K. W. (1970). *J. Mol. Biol.* **51**, 379-391.
- Yoshimori, R. (1971). Ph.D. thesis, University of California, San Francisco, 73 pp.